IOP Publishing | Institute of Physics and Engineering in Medicine Physiological Measurement

Physiol. Meas. **37** (2016) 1370–1382 doi:10.1088/0967-3334/37/8/1370

Assessing ECG signal quality indices to discriminate ECGs with artefacts from pathologically different arrhythmic ECGs

C Daluwatte[1], L Johannesen[1], L Galeotti[1,2], J Vicente[1,3], D G Strauss[1] and C G Scully[1]

[1] Office of Science and Engineering Laboratories, CDRH, US FDA, Silver Spring, MD, USA
[2] Office of Device Evaluation, Center for Devices and Radiological Health, US FDA, Silver Spring, MD, USA
[3] BSICoS Group, Aragón Institute of Engineering Research (I3A), IIS Aragón, University of Zaragoza, Zaragoza, Spain

E-mail: Chathuri.Daluwatte@fda.hhs.gov

Received 10 February 2016, revised 26 April 2016
Accepted for publication 9 May 2016
Published 25 July 2016

Abstract

False and non-actionable alarms in critical care can be reduced by developing algorithms which assess the trueness of an arrhythmia alarm from a bedside monitor. Computational approaches that automatically identify artefacts in ECG signals are an important branch of physiological signal processing which tries to address this issue. Signal quality indices (SQIs) derived considering differences between artefacts which occur in ECG signals and normal QRS morphology have the potential to discriminate pathologically different arrhythmic ECG segments as artefacts. Using ECG signals from the PhysioNet/Computing in Cardiology Challenge 2015 training set, we studied previously reported ECG SQIs in the scientific literature to differentiate ECG segments with artefacts from arrhythmic ECG segments. We found that the ability of SQIs to discriminate between ECG artefacts and arrhythmic ECG varies based on arrhythmia type since the pathology of each arrhythmic ECG waveform is different. Therefore, to reduce the risk of SQIs classifying arrhythmic events as noise it is important to validate and test SQIs with databases that include arrhythmias. Arrhythmia specific SQIs may also minimize the risk of misclassifying arrhythmic events as noise.

Keywords: signal quality, ECG, arrhythmia

(Some figures may appear in colour only in the online journal)

1. Introduction

False patient monitor alarms induced by noise and signal artefacts occur regularly in the intensive care unit (ICU) and contribute to the high numbers of medical device alarms (Borowski *et al* 2011). This disrupts patient care by increasing stress and reducing sleep, and makes caregivers desensitized to alarms to the extent that they may miss critical clinical events (Schmid *et al* 2013). Recent studies have shown delayed caregiver reaction times to actionable alarms for patients that have higher numbers of non-actionable alarms (Bonafide *et al* 2015). These effects from high false alarm rates in the clinic are referred to as alarm fatigue and make improper medical device alarm responses one of the top health technology hazards (ECRI Institute 2015).

Patient monitors utilize sensors to record physiological signals from patients and then apply algorithms to produce physiological measurements, monitor the patient's state, and notify clinical staff of critical events by triggering alarms. In a recent single hospital study, 88.8% of arrhythmia alarms produced by patient monitors in the ICU were shown to be false or non-actionable (Drew *et al* 2014). High numbers of non-actionable alarms produced by patient monitors may be caused by inappropriate alarm thresholds for the particular patient (e.g. utilizing a low SpO_2 threshold alarm of 90% when the patient's average SpO_2 is 88%) or artefacts in the signals resulting from poor sensor-patient connections during periods of motion (Borowski *et al* 2011).

Significant effort has been made in the field of physiological signal processing to develop computational approaches to automatically identify artefacts in ECG signals. For example, the Computing in Cardiology 2011 challenge focused specifically on identifying low quality ECGs from mobile recorders to inform if a new recording should be taken (Clifford *et al* 2012). Such methods, in general, first quantify a feature of the ECG signal that is expected to be related to the amount of signal artefact disruptive to further computation of the measure of interest and then apply a threshold to an individual feature or classifier to multiple features. These features are referred to as signal quality indices (SQIs). SQIs are often developed and tested using ECG waveform datasets annotated as clinically usable if a human expert can derive a reliable heart rate and unusable otherwise (Clifford and Moody 2012, Orphanidou *et al* 2015). While such datasets provide a basis for motion and other artefacts in ECG waveforms, they often lack an important subset of ECG waveforms present in the clinical environment: pathologically different arrhythmic ECG waveforms which may be mistaken as noise. Arrhythmic ECGs contain different signal characteristics than normal ECGs, and different arrhythmias present different patterns on ECGs. Therefore, it is unknown how SQIs perform in distinguishing signal artefacts from pathological arrhythmias for specific arrhythmic ECG waveforms. There is a risk that an arrhythmic ECG may be misclassified as a noisy recording, thus preventing the detection of clinically relevant events.

In this study, we assessed if SQIs previously reported in the scientific literature can differentiate ECG segments with artefacts from arrhythmic ECG segments. We annotated the quality of the ECG signals during arrhythmia alarms (asystole, bradycardia, and tachycardia) in The PhysioNet/Computing in Cardiology Challenge 2015 training set (Clifford *et al* 2015). We characterized the distributions of SQIs to assess: (1) if previously reported SQIs can distinguish heart-rate based alarms induced by signal artefacts from those induced by arrhythmia patterns and (2) what the most discriminative SQIs are irrespective of the arrhythmia.

2. Methods

2.1. Experimental dataset

We used data from The PhysioNet/Computing in Cardiology Challenge (CinC) 2015 training set. Full details on this dataset are provided in (Clifford *et al* 2015). Briefly, the dataset consisted of at least 5 min records with two ECG signals and photoplethysmogram and/or blood pressure waveform recordings from bedside monitors. In each record an arrhythmia alarm (asystole: no heartbeats for 4 s, extreme bradycardia: heart rate lower than 40 beats per minute (bpm), extreme tachycardia: heart rate higher than 140 bpm, ventricular tachycardia: 5 or more consecutive ventricular beats within 2.4 s (a rate of 100 bpm), ventricular fibrillation/flutter: rapid fibrillatory, flutter, or oscillatory waveform for 4 s) was triggered by the bedside monitor at the 5 min mark, with the onset of the event occurring at some point in the preceding 10 s. The type of alarm detected is indicated in each record. Expert human annotators reviewed and labelled each alarm as either true or false using data from all waveforms recorded from the patient before and after the alarm event.

2.2. Labelling of ECG quality

For our analysis on ECG SQIs we used the ECG signals in the CinC 2015 training set. Ventricular tachycardia and ventricular fibrillation/flutter result in extreme modification of the ECG waveforms, and it is unclear if the SQI techniques are appropriate to discriminate these arrhythmias from artefacts, thus we focused our analysis on asystole, bradycardia, and tachycardia alarms. For these alarms, we annotated ECG signals during the 10 s alarm period prior to alarm trigger (4:50–5:00) as 'high quality' or 'low quality' signals. The dataset was randomly divided into three sets, and each set was assigned to two out of three reviewers (experienced researchers in cardiovascular signal processing) in a way that each random set has a unique two reviewer combination (reviewer 1 and 2, reviewer 1 and 3 or reviewer 2 and 3). The two reviewers independently labelled the last 10 s segment of each ECG signal, the period including the event that triggered the alarm, by viewing the waveform segment with marked locations of QRS detections. If the reviewer considered that the heartbeat detector did not identify correct heartbeat locations in the segment due to artefacts it was labelled as low quality. A set of high/low quality reference annotations was generated from the segments where both reviewers agreed on high/low quality labelling. Segments that two reviewers did not agree on were not considered in the analysis. The distribution of high/low quality labelling and disagreement is reported in table 1.

During review of the asystole records, some of the signals were identified to contain pacemaker pulses which appeared distorted (likely due to the low sampling frequency of the dataset). Considering that bedside monitors usually have pre-processing stages specific for detection and removal of pacemaker pulses that may handle the presence of distorted pacemaker pulses differently than other artefacts triggering alarms, we excluded these records from further analysis. This resulted in 680 total ECG signals used from the dataset.

2.3. Selection of signal quality indices

A number of SQIs which classify ECG signals as either valid or corrupt were identified after a literature search on PubMed for ((ECG OR electrocardiogram) AND (signal quality OR artefact detection)). SQIs producing a continuous quantitative measure related to the quality of the signal that would further be used in a classification system were identified. A subset of these

Table 1. Labelling of ECG signal quality for asystole, bradycardia, tachycardia alarms.

	Asystole		Bradycardia		Tachycardia	
	True alarm	False alarm	True alarm	False alarm	True alarm	False alarm
High quality	24	70	76	47	230	9
Low quality	8	94	2	32	6	7
Disagreement	2	24	14	7	26	2
Total	34	188	92	86	262	18

SQIs was then selected based on the information available to implement the computational method on 10 s single-lead ECG signals. The list of SQIs selected for this study is summarized in table 2. In the original literature, 10 s epochs were used to calculate each of these SQIs from the ECG. For visualization purposes we added the thresholds derived in the original literature for each of these SQIs to the figures.

2.4. Applying signal quality indices to arrhythmia data

Signal analysis was performed in Matlab R2014b (The Mathworks, Natick, MA). SQIs listed in table 2 were calculated from the ECG signals using the last 10 s of the record where the event triggering the arrhythmia alarm is present. We used a previously reported QRS detector based on the U3 transform of the signal for SQIs requiring QRS or R-peak locations (Marchesi and Paoletti 2004, Paoletti and Marchesi 2006, Johannesen *et al* 2013). We characterized each SQI distribution for ECG segments with artefacts compared to arrhythmic ECG segments. For the signal quality labelled asystole, bradycardia, and tachycardia alarms, we compared the distribution of each SQI calculated using ECG segments from two groups: low quality signals from false alarms (corresponds to ECG segments with artefacts, which will be called 'ECG artefacts' hereafter in the paper) and high quality signals from true alarms (corresponds to arrhythmic ECG segments, which will be called 'arrhythmic ECG' hereafter in the paper). These two groups correspond to ECG signals that likely contain either (a) artefacts inducing false alarms or (b) arrhythmia patterns inducing true alarms. It is important to note that the true/false alarm reviewers may have based their classification on additional signals or prior sections of the record, which may explain some of the false alarms.

The criteria to trigger the different arrhythmia alarms may also cause variability in the SQI distributions among arrhythmic ECGs. Therefore, we further studied the ability of SQIs to discriminate between 'ECG artefacts' and 'arrhythmic ECG' independently for asystole, bradycardia and tachycardia to assess if the same SQIs are applicable to each alarm type.

2.5. Statistical analysis

The ability of each SQI to discriminate 'ECG artefacts' and 'arrhythmic ECG' was assessed with area under the receiver operating characteristic curve (AUC) using package pROC (Robin *et al* 2011) in R version 3.2.2 (R Foundation for Statistical Computing, Vienna, Austria). We identified the three best SQIs in terms of their discriminative potential using the three highest AUC values. For each record there were two ECGs labelled as high quality, low quality, or unknown. Because this quality assessment was made for each ECG lead, and not for the record, we considered each ECG signal independently.

Table 2. Implemented signal quality indices.

Paper	Features	Variable name
Orphanidou *et al* (2015)	Mean heart rate for 10 s	*meanhr*
	Maximum RR interval for 10 s	*maxrri*
	Maximum RR interval/Minimum RR interval for 10 s	*maxrr2minrr*
	Average template matching correlation coefficient: average of the correlation coefficients of each QRS complex with mean QRS complex	*avecorr*
Di Marco *et al* (2012)	Low-frequency time marginal energy (0–0.5 Hz)	*mELF*
	QRS-band time marginal energy (5–25 Hz)	*mEQRS*
	High-frequency time marginal energy (>100 Hz)	*mEHF*
	Median value of the peak-to-nadir amplitude difference of the QRS complexes in 10 s	*qrsa*
Hayn *et al* (2012)	Portion of samples that are the same	*constamppct*
	Portion of samples situated close to spikes: percentage of samples that are <0.02 in the low pass filtered (0–10 Hz) binary coded signal, where the binary coded signal is obtained by coding the 30–70 Hz bandlimited signal as 1 if $\lvert sample_{i+1} - sample_i \rvert > 0.1$ mV ms^{-1}, 0 otherwise.	*close2spikepct*
	QRS detection based parameter: normalized difference between the amplitude of the smallest R peak and the highest amplitude of non-QRS peak, with respect to the highest amplitude of non-QRS peak. R peaks are the maxima which reside on to the left of the most distinct step in the descending ordered maxima sequence while the non-QRS peaks are the ones to the right.	Q_{rel}
	Ratio of the standard deviation of RR interval to mean RR interval	*sdrr2meanrr*
	Range of signal amplitude around QRS detection: maximum minus the minimum signal amplitude within a QRS complex	*rangeqrs*
Li *et al* (2008)	Kurtosis of the signal for 10 s ECG segment	*kur*
	Ratio of the sum of the power of the ECG between frequencies of 5–14 Hz to the power between 5–50 Hz	*powratio*
Johannesen *et al* (2012)	Baseline wander estimation using cubic spline (Meyer and Keiser 1977)	*bw*
	Power line noise estimation using regression-subtraction (Bazhyna *et al* 2003, Levkov *et al* 2005)	*pln*
	Residual noise by subtracting the estimated signal (median over 10 s) after subtracting baseline wander and power line noise	*residualn*

3. Results

There were a total of 330 ECG signal segments meeting our definition of 'arrhythmic ECG' (i.e. high quality signals from true alarms) and 133 meeting our definition of 'ECG artefacts' (i.e. low quality signals from false alarms) across the three arrhythmia types. However for some signal segments, certain SQIs were incalculable (e.g. when there are no beat detections, average template matching correlation coefficient (*avecorr*) cannot be calculated), which

Table 3. AUC values for SQIs for classification between 'ECG artefacts' and 'arrhythmic ECG'. *: three highest AUC for each column.

	SQI	All arrhythmias	Asystole	Bradycardia	Tachycardia
Paper	$N_{arrhythmic}$:$N_{artefact}$	330:133	24:94	76:32	230:7
Orphanidou *et al* (2015)	*meanhr*	0.61	0.77	0.87	0.99*
	maxrri	0.67	0.60	0.61	0.92*
	maxrr2minrr	0.74	0.79	0.70	0.71
	avecorr	0.93*	0.93*	0.96*	0.99*
Di Marco *et al* (2012)	*mELF*	0.67	0.90*	0.71	0.86
	mEQRS	0.68	0.74	0.75	0.55
	mEHF	0.75	0.76	0.83	0.58
	qrsa	0.52	0.59	0.61	0.69
Hayn *et al* (2012)	*constamppct*	0.50	0.50	0.50	0.50
	close2spikepct	0.60	0.79	0.72	0.64
	Q_{rel}	0.52	0.80	0.53	0.55
	sdrr2meanrr	0.62	0.55	0.71	0.48
	rangeqrs	0.65	0.55	0.78	0.56
Li *et al* (2008)	*kur*	0.71	0.88*	0.78	0.76
	powratio	0.53	0.86	0.45	0.69
Johannesen *et al* (2012)	*bw*	0.79*	0.55	0.90*	0.64
	pln	0.79*	0.56	0.87	0.56
	residualn	0.87*	0.56	0.98*	0.82

introduced missing values in the dataset. Table 3 summarizes the AUC values for the SQI between the 'ECG artefacts' and 'arrhythmic ECG' groups for all alarm types combined and each alarm type individually. All SQIs but *constamppct* had AUC > 0.5 indicating they have some potential to classify 'arrhythmic ECG' and 'ECG artefacts'. In descending order, average template matching correlation coefficient (*avecorr*), residual noise (*residualn*), baseline wander (*bw*) and power line noise (*pln*) are the SQIs with the highest AUC when combining the ECGs from different arrhythmias (AUC $= 0.93, 0.87, 0.79$ and 0.79 respectively). The box dot plots of these four SQIs (figure 1) show that the average template matching correlation coefficient (*avecorr*) has superior discriminatory potential over the other three SQIs. Although 'ECG artefacts' and 'arrhythmic ECGs' come from different distributions for the other three SQIs, the distributions overlap considerably, limiting their potential to differentiate signals contaminated with artefacts from those generated from arrhythmias. For all four SQIs, the 'arrhythmic ECG' distributions are much narrower than those contaminated with artefacts.

When comparing discrimination ability between 'ECG artefacts' and 'arrhythmic ECG', for specific arrhythmias, the average template matching correlation coefficient (*avecorr*) still had a high AUC (AUC: 0.93–0.99, table 3) for each type of arrhythmia. This indicates that *avecorr* produces different distributions between 'ECG artefacts' and 'arrhythmic ECG' for each of the three arrhythmia patterns in the dataset. As can be seen in the first column of figure 2, asystole, bradycardia, and tachycardia arrhythmia patterns produce high correlation coefficients between each individual pulse and the template pulse generated from the mean of all pulses. This distinguishes the signals from artefacts that induce each alarm and disrupt the consistent QRS waveform morphology, lowering *avecorr*.

Besides the average template matching correlation coefficient, other SQIs are identified in each column of table 3 that may produce different distributions between signal artefacts and

specific arrhythmia patterns. The next two SQIs that have the highest AUC for each arrhythmia type are different. For asystole signals, kurtosis of the signal (*kur*) and low-frequency time marginal energy (*mELF*) each have different distributions between 'ECG artefacts' and 'arrhythmic ECG' (figure 2). Different distributions between 'ECG artefacts' and bradycardia ECG segments were observed for residual noise (*residualn*) and baseline wander (*bw*) (figure 2). As seen in the asystole column in table 3, the baseline wander (*bw*) and residual noise (*residualn*) SQI estimates have poor classification ability for 'ECG artefacts' and asystole segments. Tachycardia ECG segments have different distributions with respect to 'ECG artefacts' for the maximum RR interval for 10 s (*maxrri*) and mean heart rate range for 10 s (*meanhr*) (figure 2). It is important to note that ECG signal dataset for tachycardia alarms is highly skewed towards false alarms with a true to false alarm ratio of 230:7.

4. Discussion

Signal quality indices derived considering differences between ECG segments with artefacts and ECG segments with high quality QRS morphology have the potential to discriminate pathologically different arrhythmic ECG segments as artefacts. We studied which SQI can differentiate ECG segments with artefacts from arrhythmic ECG segments. We found that although some SQIs may differentiate artefacts from multiple pathological arrhythmia patterns, other SQIs may consider certain arrhythmia patterns as artefacts.

As reported we identified SQIs which differentiate ECG segments with artefacts from arrhythmic ECG segments using the metric AUC. We also compared distributions of two groups using Kolmogorov–Smirnov (KS) test statistic and used the significant KS statistic values ($p < 0.05$) to identify best discriminators, which resulted in the same order of SQIs as produced by AUC. While AUC is a measure of discriminative ability, KS statistic is a measure of distance/difference in shape between cumulative distribution functions. Because both approaches provided the same results, we did not report the KS test results. Based on these best discriminators machine learning algorithms/ multi-lead strategies can be implemented to reduce misclassification of alarms.

The average template matching correlation coefficient (*avecorr*) consistently differentiated 'ECG artefacts' from 'arrhythmic ECG' for asystole, bradycardia and tachycardia alarms. This SQI uses a normal beat template and compares it to each individual detected beat (Orphanidou *et al* 2015). Because asystole, bradycardia and tachycardia alarms will generally have consistent morphology between beats, the correlation between beats in the record will be high. However during periods of noise and signal artefacts falsely detected beats will not match with normal QRS morphology, producing a lower correlation coefficient. Other individual SQIs previously reported to distinguish clean from artefact-containing ECGs have the potential to classify pathologically different arrhythmic ECG segments as artefacts. For example, measures of baseline wander (*bw*) work well to discriminate bradycardia from artefacts but not asystole. This may be because the lack of beats in true asystole patterns may affect the low frequency characteristics of the signal. Therefore during algorithm development it is important to consider whether to run the same signal quality tests for all types of arrhythmia alarms, or to develop indices to assess the signal quality that considers expected patterns in the specific arrhythmia. It should also be noted that although we studied single SQIs here, to differentiate 'ECG artefacts' from 'arrhythmic ECG' multiple SQI combinations or multiple physiological signals could help improve the discrimination ability (Li and Clifford 2012). The SQIs assessed were all used in conjunction with other SQIs when implemented in the original literature reports.

For tachycardia alarms, two of the highest discriminating SQIs were the maximum RR interval for 10 s (*maxrri*) and mean heart rate for 10 s (*meanhr*). The paper describing the mean heart rate as an SQI (Orphanidou *et al* 2015) originally used thresholds of less than

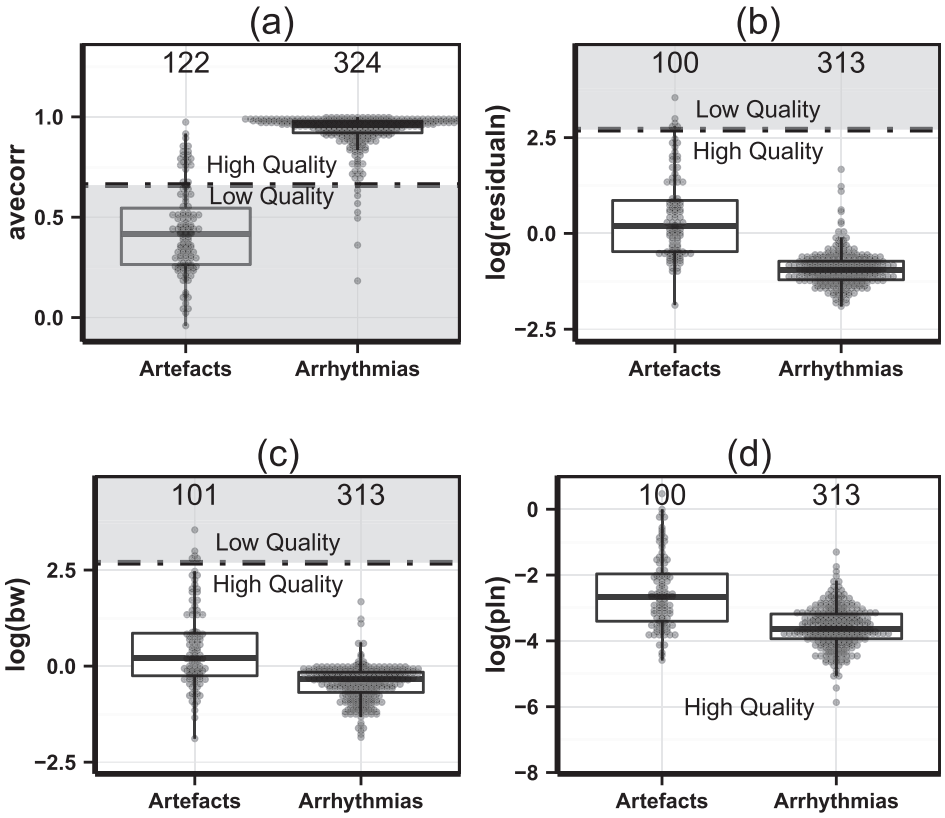

Figure 1. Box dot plots for two groups: 'ECG artefacts' and 'arrhythmic ECG' from asystole, bradycardia, and tachycardia alarms for SQI: (a) average template matching correlation coefficient (*avecorr*); (b) residual noise (*residualn*) (plotted in $\log_{10}$ scale); (c) baseline wander (*bw*) (plotted in $\log_{10}$ scale); and (d) power line noise (*pln*) (plotted in $\log_{10}$ scale, noise threshold outside the graph ($\log_{10}13\,680 = 4.1$)). The number of ECG signals in each group is shown above the boxes. Shaded regions represent the low quality region defined with respect to the threshold from original literature. In the box plot the upper and lower hinges correspond to the 25th and 75th percentiles while the middle one corresponds to the median. The whiskers correspond to the highest value within 1.5* inter quartile range. The overlaying dot plot shows the observed data divided into bins.

40 or greater than 180 bpm to classify poor signal quality. Our data was mostly within these bounds for both groups, however 4 true tachycardia ECG segments were above the 180 bpm threshold and would be classified as low quality. It is also important to note that this SQI was not used individually to assess signal quality but as one branch of a decision tree (Orphanidou *et al* 2015), likely to exclude grossly misdetected heartbeats due the presence of noise. For use in the presence of arrhythmia, the threshold for this alarm should be placed above the maximum reasonable heart rate in case of tachycardia, arguably up to 200–300 bpm according to AAMI/ANSI/IEC 60601-2–27:2011, medical electrical equipment—part 2–27: particular requirements for the basic safety and essential performance of electrocardiographic monitoring equipment. Similar considerations should be taken into account for all the SQI that rely on measurements that define the arrhythmic conditions.

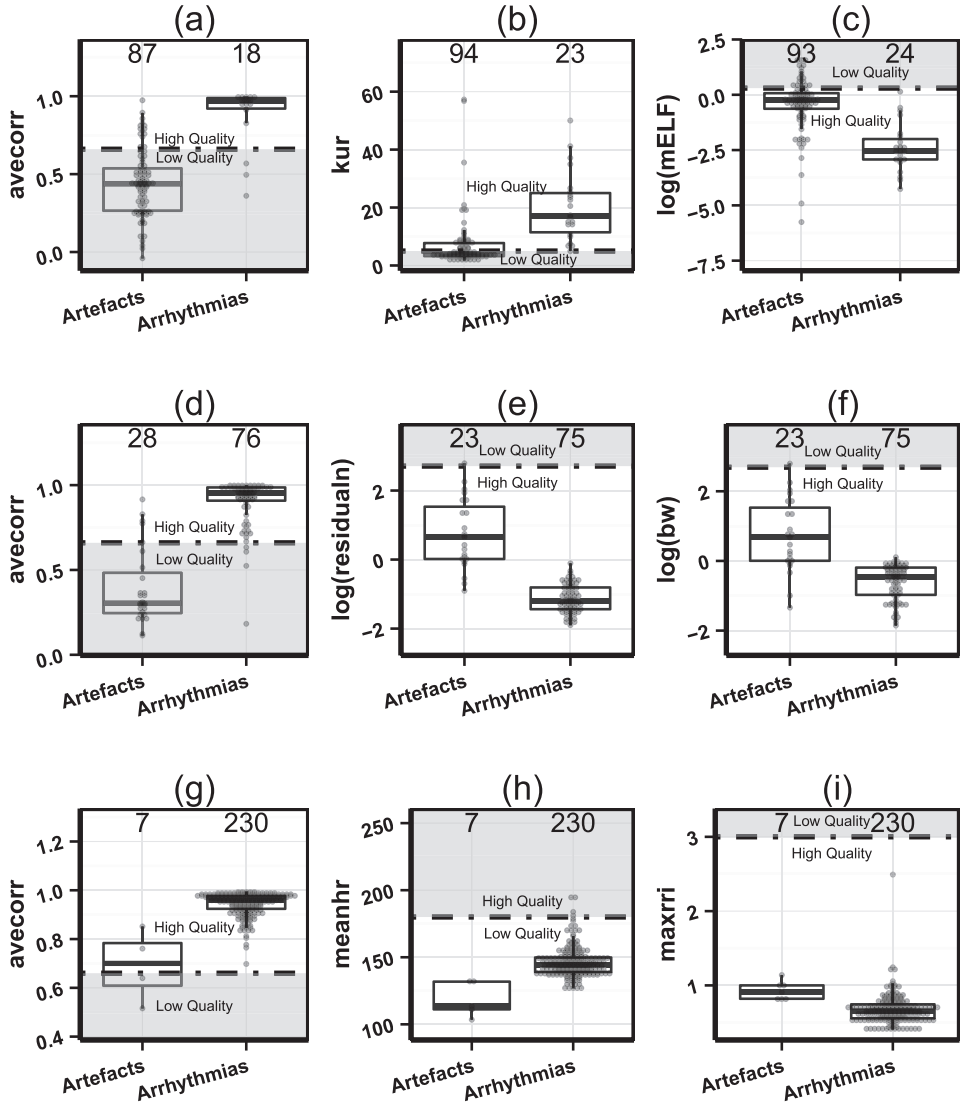

Figure 2. Box dot plots for two groups: 'ECG artefacts' and 'arrhythmic ECG' from asystole alarms for SQI: (a) average template matching correlation coefficient (*avecorr*), (b) kurtosis of the signal (*kur*) and (c) low-frequency time marginal energy (*mELF*) (plotted in $\log_{10}$ scale); from bradycardia alarms for SQI: (d) average template matching correlation coefficient (*avecorr*), (e) residual noise (*residualn*) (plotted in $\log_{10}$ scale) and (f) baseline wander (*bw*) (plotted in $\log_{10}$ scale); from tachycardia alarms for SQI: (g) average template matching correlation coefficient (*avecorr*) (h) mean heart rate range for 10 s (*meanhr*) and (i) maximum RR interval for 10 s (*maxrri*). Shaded regions represent the low quality region defined with respect to the threshold from original literature. In the box plot the upper and lower hinges correspond to the 25th and 75th percentiles while the middle one corresponds to the median. The whiskers correspond to the highest value within 1.5* inter quartile range. The overlaying dot plot shows the observed data divided into bins.

Analysis of the results helped identify potential limitation of the dataset. As an example, we found that some true tachycardia alarm had a heart rate below 140 bpm (a heart rate of greater than 140 bpm for 17 consecutive beats was the challenge definition of tachycardia), so if the heart rate measured on the bedside monitor was lower than the tachycardia threshold the alarm would not have been triggered. This suggests one of at least three scenarios: (1) the tachycardia alarm was triggered by a different signal, (2) the tachycardia threshold limit on the bedside monitor was not the same as in the challenge definitions, or (3) the beat locations identified by our QRS detector are different than the beat locations identified by the bedside monitor. Considering the latter, the distributions from the original QRS detections which triggered the alarm might have a different distribution than from the QRS detector we used. All the SQI based on QRS detections are dependent of the specific QRS detector used, thus different detectors may lead to different performances. Different QRS detectors may fail at different noise conditions in the signal so by combining detections from multiple QRS detectors the robustness of heart rate measurements has been shown to be improved (Liu *et al* 2013). This concept of using multiple QRS detectors and merging their beat location estimates can be extended to improve the robustness of SQIs for specific arrhythmia types (Li *et al* 2008).

From our low/high quality annotations we found that for asystole, bradycardia and tachycardia alarms in this dataset those annotated as true were likely to be generated from high quality signals while those annotated as false were a mix of high and low quality signals. For our analysis we only used low quality segments from false alarms and high quality segments from true alarms to focus on the groups we believed contained ECG with artefacts and ECG with arrhythmias. True alarms considered low quality likely indicates that the alarm was triggered and verified by the human annotator from a different signal. False alarms considered high quality likely indicates that the alarm was triggered from a different signal but may have been verified as false using the current signal. In either case, these were not likely to have been the signals that would have triggered the alarm, but may be useful for checking alarm conditions in multi-parameter monitoring systems. We looked at the SQI distributions for low quality segments from true alarms and high quality segments from false alarms to understand where these distributions reside. We found that for the best discriminator, average template matching correlation coefficient (*avecorr*), low quality segments from true alarms overlaps with low quality segments from false alarms, and high quality segments from false alarms overlaps with high quality segments from true alarms (figure 3(a)). However for other SQIs, this was not always true (figures 3(b)–(d)). Humans looking at ECGs might classify ECG as low quality if (1) the QRS complex is not consistent in morphology, (2) there is a high amount of baseline wander and/or (3) significant power line noise is present. Therefore there may be an initial bias towards these SQIs' superior potential to discriminate low quality signals annotated by humans.

For visualization purposes we added the thresholds derived in original literature for each of these SQIs to the figures. However it is important to note that these SQIs and corresponding thresholds were not used individually but in combination with other SQIs in the original literature sources. Another factor that may affect the performance of SQIs in our current study compared to in their original sources is the data the SQIs were tested on. The databases in the original literature were recorded with high sampling rates as opposed to the current CinC dataset which was provided at 250 Hz. This could result in different differentiating thresholds.

The CinC 2015 training set provided a unique opportunity to look at performance of SQIs for specific alarm types (Clifford *et al* 2015). Utilizing such data will likely augment other types of databases for assessing signal quality such as the use of motion studies in otherwise normal ECG. Although many SQIs have been developed with otherwise normal data and shown to sufficiently discriminate periods of ECG with artefacts from clean QRS morphology,

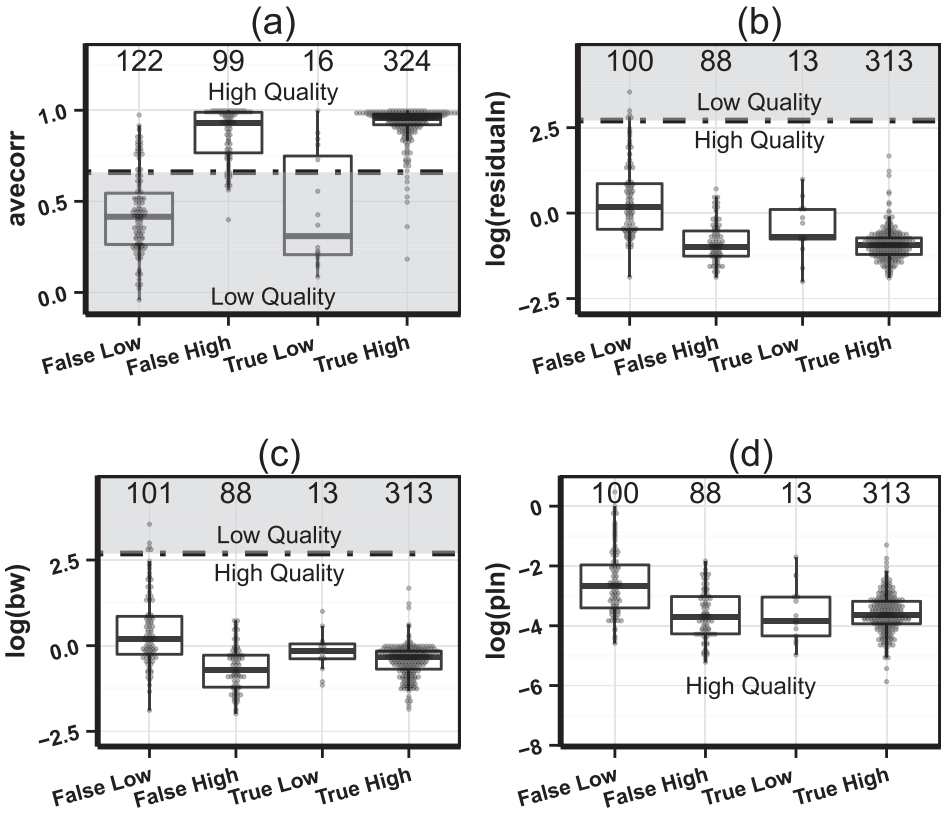

Figure 3. Box dot plots for four groups: 'low quality signals from false alarms', 'high quality signals from false alarms', 'low quality signals from true alarms' and 'high quality signals from true alarms' from asystole, bradycardia, and tachycardia alarms for SQI: (a) average template matching correlation coefficient (*avecorr*); (b) residual noise (*residualn*) (plotted in $\log_{10}$ scale); (c) baseline wander (*bw*) (plotted in $\log_{10}$ scale); and (d) power line noise (*pln*) (plotted in $\log_{10}$ scale, noise threshold outside the graph ($\log_{10}13680 = 4.1$)). Shaded region represent the low quality region defined with respect to the threshold from original literature. In the box plot the upper and lower hinges correspond to the 25th and 75th percentiles while the middle one corresponds to the median. The whiskers correspond to the highest value within 1.5* inter quartile range. The overlaying dot plot shows the observed data divided into bins.

testing on clinical datasets with labelled arrhythmias enables a better understanding of performance in clinical settings (Orphanidou *et al* 2015). It also allows identification of signal parameters under specific arrhythmia conditions and how they relate to noise. This could lead to a better understanding of the characteristics of noise and arrhythmia patterns for the development of more robust noise stress tests for new ECG processing algorithms (Moody *et al* 1984).

Ventricular tachycardia and ventricular fibrillation alarms were also present in the CinC 2015 training set. We did not consider these alarm types in the present study. Most of the SQIs listed in table 2 were originally developed to discriminate ECG segments with artefacts from ECG segments with high quality QRS morphology. SQIs assessing kurtosis of the signal for 10 s ECG segments and ratio of the sum of the power of the ECG between frequencies of 5–14 Hz

to the power between 5–50 Hz were developed and used in heart rate based arrhythmic signals (Li *et al* 2008). In addition, the QRS detector is designed to detect normal and non-tachycardia ventricular beats, however it may be unreliable to detect beats during periods of ventricular tachycardia and ventricular fibrillation/flutter. Therefore we looked at ECG segments from heart rate based arrhythmias (asystole, bradycardia and tachycardia) while excluding those where the QRS morphology is pathologically different.

5. Conclusion

There is no one-size-fits all SQI when considering arrhythmias. Since the pathology of different arrhythmias can produce different patterns in the ECG waveforms it is likely that the distribution of SQIs will vary based on the arrhythmia type. We found this to be true in our study as the best SQI changed with the arrhythmia type. Considering specific SQIs based on the arrhythmia type may minimize the risk of classifying arrhythmic events as noise. It will not always be true that SQIs developed from normal rhythms will be applicable to pathological ECGs. Therefore testing SQIs with databases of specific arrhythmia types can improve the design of SQIs and robust monitoring systems.

Disclaimer

The mention of commercial products, their sources, or their use in connection with material reported herein is not to be construed as either an actual or implied endorsement of such products by the Department of Health and Human Services.

Acknowledgments

This project was supported in part by U.S. Food and Drug Administration's Medical Countermeasures Initiative, Critical Path Initiative, Office of Women's Health and appointments to the Research Participation Programs at the Oak Ridge Institute for Science and Education through an interagency agreement between the Department of Energy and FDA.

References

Bazhyna A, Christov I, Gotchev A, Daskalov I and Egiazarian K 2003 Powerline interference suppression in high-resolution ECG *Computers in Cardiology* (IEEE) pp 561–4

Bonafide C P *et al* 2015 Association between exposure to nonactionable physiologic monitor alarms and response time in a children's hospital *J. Hosp. Med.* **10** 345–51

Borowski M, Gorges M, Fried R, Such O, Wrede C and Imhoff M 2011 Medical device alarms *Biomed. Tech.* **56** 73–83

Clifford G D and Moody G B 2012 Signal quality in cardiorespiratory monitoring *Physiol. Meas.* **33**

Clifford G D, Silva I, Moody B, Li Q, Kella D, Shahin A, Kooistra T, Perry D and Mark R G 2015 The PhysioNet/Computing in cardiology challenge 2015: reducing false arrhythmia alarms in the ICU *Computing in Cardiology Conf. (Nice, 6–9 Septemeber 2015)* pp 273–6

Di Marco L Y, Duan W, Bojarnejad M, Zheng D, King S, Murray A and Langley P 2012 Evaluation of an algorithm based on single-condition decision rules for binary classification of 12-lead ambulatory ECG recording quality *Physiol. Meas.* **33** 1435–48

Drew B J *et al* 2014 Insights into the problem of alarm fatigue with physiologic monitor devices: a comprehensive observational study of consecutive intensive care unit patients *PLoS One* **9** e110274

ECRI Institute 2015 Top 10 health technology hazards for 2016 (www.ecri.org/Resources/Whitepapers_and_reports/2016_Top_10_Hazards_Executive_Brief.pdf) (accessed 12 November 2015)

Hayn D, Jammerbund B and Schreier G 2012 QRS detection based ECG quality assessment *Physiol. Meas.* **33** 1449–61

Johannesen L and Galeotti L 2012 Automatic ECG quality scoring methodology: mimicking human annotators *Physiol. Meas.* **33** 1479–89

Johannesen L, Vicente J, Galeotti L and Strauss D G 2013 Ecglib: library for processing electrocardiograms *Computing in Cardiology (IEEE)* pp 951–4

Levkov C, Mihov G, Ivanov R, Daskalov I, Christov I and Dotsinsky I 2005 Removal of power-line interference from the ECG: a review of the subtraction procedure *Biomed. Eng. Online* **4** 50

Li Q and Clifford G D 2012 Signal quality and data fusion for false alarm reduction in the intensive care unit *J. Electrocardiol.* **45** 596–603

Li Q, Mark R G and Clifford G D 2008 Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a Kalman filter *Physiol. Meas.* **29** 15–32

Liu N T, Batchinsky A I, Cancio L C, Baker W L Jr and Salinas J 2013 Development and validation of a novel fusion algorithm for continuous, accurate, and automated R-wave detection and calculation of signal-derived metrics *J. Crit. Care* **28** 885 e9–18

Marchesi C and Paoletti M 2004 ECG processing algorithms for portable monitoring units *Internet J. Med. Technol.* **1** 2

Meyer C R and Keiser H N 1977 Electrocardiogram baseline noise estimation and removal using cubic splines and state-space computation techniques *Comput. Biomed. Res.* **10** 459–70

Moody G B, Muldrow W and Mark R G 1984 A noise stress test for arrhythmia detectors *Comput. Cardiol.* **11** 381–4

Orphanidou C, Bonnici T, Charlton P, Clifton D, Vallance D and Tarassenko L 2015 Signal-quality indices for the electrocardiogram and photoplethysmogram: derivation and applications to wireless monitoring *IEEE J. Biomed. Health Inform.* **19** 832–8

Paoletti M and Marchesi C 2006 Discovering dangerous patterns in long-term ambulatory ECG recordings using a fast QRS detection algorithm and explorative data analysis *Comput. Methods Prog. Biomed.* **82** 20–30

Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C and Müller M 2011 pROC: an open-source package for R and S+ to analyze and compare ROC curves *BMC Bioinform.* **12** 1–8

Schmid F, Goepfert M S and Reuter D A 2013 Patient monitoring alarms in the ICU and in the operating room *Crit. Care* **17** 216 (PMID: 23510435)